论文参考文献学科多样性与被引频次的关系研究*

杨 雪

中山大学信息管理学院 广州 510006

摘 要 [目的/意义] 构建论文参考文献学科多样性测度指标，探究论文参考文献学科多样性与被引频次之间的关系。**[方法/过程]** 基于 Rao-Stirling 多样性指数设计参考文献学科多样性测度指标，并以论文参考文献学科多样性为自变量，论文被引频次为因变量，期刊等级等 15 个变量为控制变量，使用普通最小二乘法拟合生成回归模型，探讨论文参考文献学科多样性与被引频次之间的关系。**[结果/结论]** 研究构建的参考文献多样性测度指标能够反映论文的多学科知识融合程度；论文参考文献学科多样性正向影响论文被引频次，论文参考文献学科多样性越强，论文被引频次越高。

关键词：科技评价 参考文献学科多样性 交叉科学 学术影响力

分类号：G301

被引频次是论文受学界关注程度和论文自身学术影响力的直观体现，其在一定程度上能够反映论文质量与学术价值[1-2]。基于论文被引频次的相关评价指标已成为当前各机构开展科研绩效评价和科研人员职位晋升的主要依据，如 h 指数等。明确论文被引频次的作用规律、影响因素有利于为科学管理工作提供参考，同时也能够进一步明确科研成果的传播机制。现有研究探讨了作者学术经验[3-4]、合作模式[5]、论文来源地理位置[6]、标题形式[7]、论文长度[8]以及是否开放获取[9]等因素与论文被引频次之间的影响关系，但对论文参考文献与被引频次之间关系的研究依旧留有探索空间。

参考文献是论文的知识基础，从论文知识生产过程来看，科学知识生产具有关联性和累积性，知识基础与论文产出内容属于同一体系，几乎所有知识都是相关科研主体在已建构的知识基础上进行的延伸[10-12]，知识基础直接作用于论文内容并进一步对其后续传播产生影响。以参考文献为代表的知识基础能够充分体现论文研究起点以及研究者前期工作深度，是反映与影响论文质量的重要方面[13-15]。从理论上说，来自多学科的参考文献可促进融合不同领域知识，激发新的

[作者简介] 杨雪，中山大学信息管理学院硕士研究生，Email：yangxue2115@163.com。

研究思路，也会吸引来自更多学科的引用[2]，能够提高论文可见度与影响力。基于现有研究对论文参考文献学科多样性与被引频次之间关系探讨不足的研究现状，本文在构建指标测度论文参考文献学科多样性的基础上，采用回归分析的方法研究论文参考文献学科多样性与被引频次之间的影响关系，以期能够为把握论文引用规律、为相关管理部门制定科技发展政策提供一定参考。

1 研究方法与研究设计/ Research Methodology and Research Design

1.1 参考文献学科多样性测度方法

参考文献学科多样性体现了论文知识基础的学科交叉程度，学科多样性强调学科之间的异质性，因此在测度参考文献学科多样性时，不仅要数学科数量，也要对参考文献来源学科之间的差异度进行衡量。学科多样性测度的前提是选取合适的学科分类体系，本文以在学界认可度较高的 WoS 学科分类体系映射参考文献的来源学科。在多样性测度指标上，目前学界已有信息熵（Shannon Entropy）、布里渊指数（Brillouin's Index）、辛普森多样性指数（Simpson Index）等多种测度指标，但上述指标大多从 1 个或 2 个维度衡量且计算复杂，尚未形成系统化的量化体系[16]。现有研究中，Stirling 基于构建的涵盖丰富度、差异度及均衡度的多样性测度指数能够充分反映系统要素分布的多样性特征，相较其他指标在综合性方面表现出极大优势[17-18]，现已获得学界的广泛认可与应用[19]，因此本文使用该指数计算论文参考文献学科多样性。依据该指数，本文构建的参考文献学科多样性测度指数包含三个基本维度，如图 1 所示。

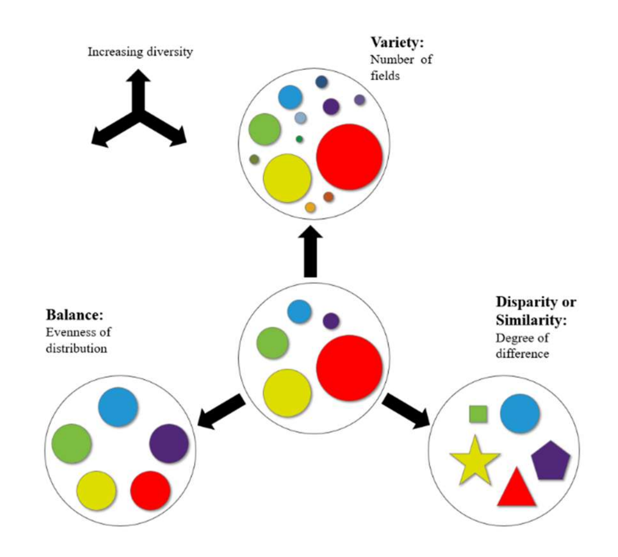


图 1 学科多样性测度维度

Figure 1 Dimensions of the disciplinary diversity measure

在本文中，Variety 指参考文献来源学科数量，Balance 指参考文献来源学科分布的均衡性，Disparity 指参考文献来源学科之间的差异性。参考文献所属学科数量越多、分布越均衡、差异越大，则其学科多样性越强。参考已有研究[20]，本文设计如下规则计算学科差异性：

$$D_{ij} = \begin{cases} 1，不同学科大类 \\ 0.66，同一大类不同学科 \\ 0，同一学科 \end{cases} \quad (1)$$

D_{ij}指参考文献学科差异性，若参考文献来源于不同学科大类，如 SCI 和 SSCI，则认为其多学科性最强，学科差异赋值为 1；若来源于同一学科大类下的不同学科，学科差异赋值为 0.66；若来源于同一学科，学科差异赋值为 0。综合以上，参考文献学科多样性指数的计算见公式 2。

$$PDI = \sum_{ij(i \neq j)} V_i \cdot B_j \cdot D_{ij} \quad (2)$$

公式 2 中，PDI 为参考文献的学科多样性，V_i为学科数量，B_j为学科分布均衡性，D_{ij}为学科差异性。

1.2 变量选取与测度

1.2.1 自变量与因变量

本研究的自变量为论文参考文献学科多样性，以上述构建的学科多样性测度指标测度。因变量为论文被引频次，以论文自发表至 2024 年 1 月的 5 年累计被引频次表示。

1.2.2 控制变量

（1）参考文献数量。参考文献数量能够反映研究者对相关领域的了解程度和对专业知识的运用情况[13]，是判断论文知识吸收与利用能力[21]和论文总体质量的重要参考指标[22]，已有研究表明，其能够对论文被引频次产生影响[23]，因此将参考文献数量作为控制变量，具体以原始数据中的 Cited Reference Count 表示。

（2）参考文献创新性。研究表明，具有更前沿的理论与技术的知识基础更容易激发具有创新性的知识生长点[24]，因而参考文献创新性也直接体现论文新颖度[25-26]，从而会对论文的被引频次产生影响。衡量参考文献创新性时，为避免极值等异常情况对测度结果的影响，本文以单篇论文所有参考文献创新性的中值表示论文参考文献整体创新性[27]。

（3）论文创新性

研究表明，科研论文创新性是影响论文被引频次的关键因素[28]，本文使用Wu等基于施引文献结构设计的D指数测度论文创新性[29]，并将其作为控制变量纳入回归模型。该指数的计算方法见公式3。

$$D=p_i-p_j=\frac{n_i-n_j}{n_i+n_j+n_k} \quad (3)$$

公式3中，i表示只引用被测论文，j表示既引用被测论文又引用其参考文献，k表示仅引用被测论文参考文献。D的取值范围为[-1,1]，D=1说明论文实现了突破式创新，D=0说明论文为中性突破性，D=-1说明论文未实现突破，是对已有研究的巩固[29]。

为便于进行回归分析与结果比较，本文将计算得到的创新性值皆进行有量纲的归一化处理，使原始数据落在区间[0,1]中，计算方法见公式4。

$$x_{new}=\frac{x-x_{min}}{x_{max}-x_{min}} \quad (4)$$

公式4中，x_{new}为归一化值，x为原始数值，x_{min}和x_{max}分别为原始数值中的最小值和最大值。

（4）期刊等级分数。通常而言，论文发表期刊等级越高，论文越有可能获得更多的关注。对于期刊等级分数，参考已有研究[30]，本文依据期刊分区对其进行分级赋分，赋分规则见公式5。其中S_j为期刊等级分数，最高为10分，最低为2分，若论文来源于Q1区期刊，则期刊等级赋予最高分10分，若来源于其他期刊则赋值为2分。

$$S_j=\begin{cases}10，Q1\\8，Q2\\6，Q3\\4，Q4\\2，其他\end{cases} \quad (5)$$

（5）其他控制变量。除上述变量外，参考学界研究，本文将作者数、基金数、标题长度等会也会对论文被引频次产生影响的变量作为控制变量同时纳入回归模型，各变量及其描述见表1。

表 1 变量描述

Table 1 Description of variables

变量分类	变量名称	变量描述	变量类型
因变量	PC5	论文 5 年累积被引频次	计数变量
自变量	PDI	参考文献学科多样性	连续变量
控制变量	PD	论文创新性	
	CD	参考文献创新性	
	RN	参考文献数量	计数变量
	AN	作者数	
	FN	基金数	
	PCN	学科数	
	TL	标题长度	
	KN	关键词数	
	AL	摘要长度	
	PN	页数	
	QS	期刊等级分数	分类变量
	OA	是否 OA 发表	虚拟变量
	AS	是否为作者独著论文	
	IS	是否为机构独著论文	
	CS	是否为国家独著论文	

2 实证分析/Empirical analysis

2.1 数据采集与处理

本文所用数据来源于 Computer Science，Artificial Intelligence、Engineering，Biomedical 和 Genetics & Heredity 领域，原因在于学科交叉融合往往被视为创新成果产出的动力之源[31]，而上述领域易产生突破性成果，其参考文献可能具有典型交叉性，因此选取上述领域进行研究具有一定代表性。由于本文中对论文及参考文献创新性指数的计算时间为 2022 年[32]，该指数的计算依赖论文施引结构，而通常论文的引用在其发表 2 年后能够在领域内达到较为稳定的状态，为保证论文及参考文献创新性测度结果的稳定性[33]，本文选取 2019 年的论文作为实证样本。具体而言，在 WoS 核心合集中按照领域检索的方式进行检索，将文献类型限定为 ARTICLE 和 PROCEEDINGS PAPER，最终得到论文 157,831 篇，提取的数据项包括 Article Title、Authors、Source Title、Affiliations、Cited References、Times Cited 等。

本文使用 DOI 号匹配论文 D 指数并衡量论文创新性[32]，因此，首先删除原始论文中无 DOI 号、参考文献与施引文献数量为 0 以及无突破性分数的论文，

剩余论文 110,295 篇。其次，对参考文献数据进行清洗与处理，提取所有参考文献的 DOI 号与来源期刊，构建“论文-参考文献 DOI 号”数据集、“论文-参考文献来源期刊”数据集，随后综合利用 JCR 和 ESI 期刊列表中的 21,576 本期刊信息构建参考文献来源期刊及其所属学科间的映射关系[34]。借鉴已有研究[20]，对于学科标识为“Multiple”的期刊，本文取第一学科作为期刊所属学科，最终构建“论文-参考文献所属学科”数据集。由于参考文献存在“No title captured”以及无 DOI 号等数据缺失情况，为保证研究结果的可靠性，本文计算了参考文献创新性指数及学科与原始参考文献数的实际匹配率，以二者匹配率同时达到 90%及以上的 12,949 篇论文作为最终样本，其中包含参考文献 611,880 篇。

2.2 总体结果描述性分析

首先，本文对论文被引频次的整体分布情况进行了分析，分析结果见图 2。图中横坐标为论文被引频次区间，纵坐标为论文落在特定区间中的概率，高度越高，表明总体样本中落在该区间的数据点越多。由图 2 可见，样本论文被引频次落在区间 [1,201] 中的概率最大，也就是说论文被引频次集中分布在该区间。

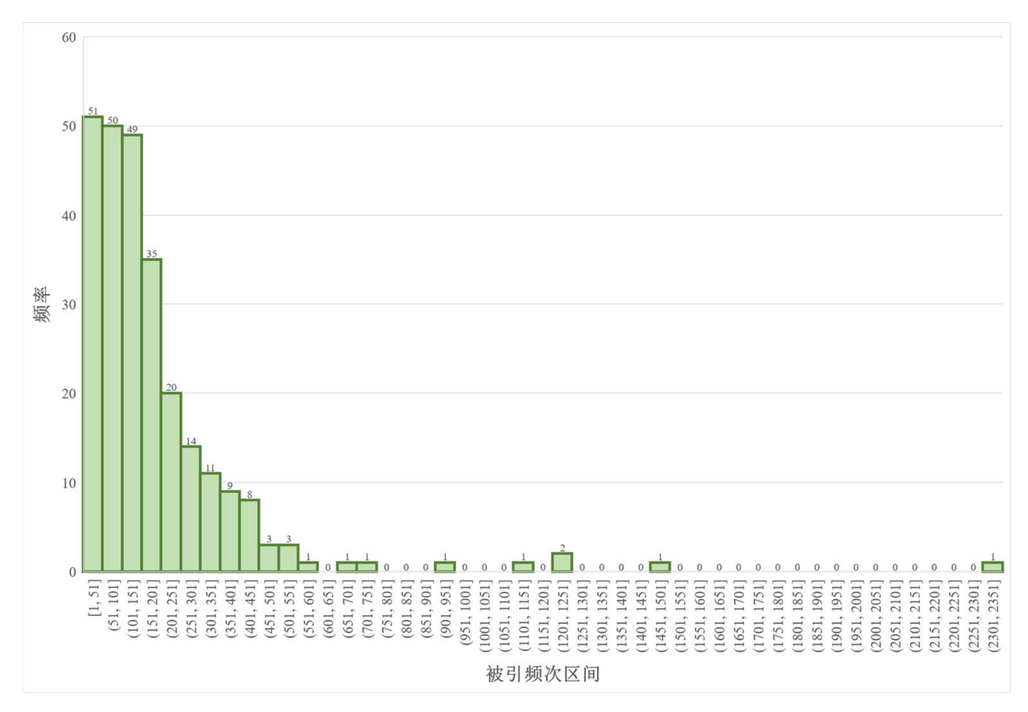


图 2 论文被引频次分布

Figure 2 Distribution of citation frequency

其次，本文对论文参考文献学科多样性的得分分布情况进行了分析，分析结果见图 3，图 3 中，横坐标为参考文献学科多样性得分区间，纵坐标为分布于特

定区间中的论文数量。由图可见，论文参考文献学科多样性集中分布于区间（0.35,0.4]和区间（0.4，0.45]中，分布于区间（0.4，0.45]中的论文数量最多，共有 6578 篇，占样本论文的 50.799%。分布于参考文献学科多样性最强的区间(0.45，0.5]中的论文有 1128 篇，占样本论文的 8.711%。

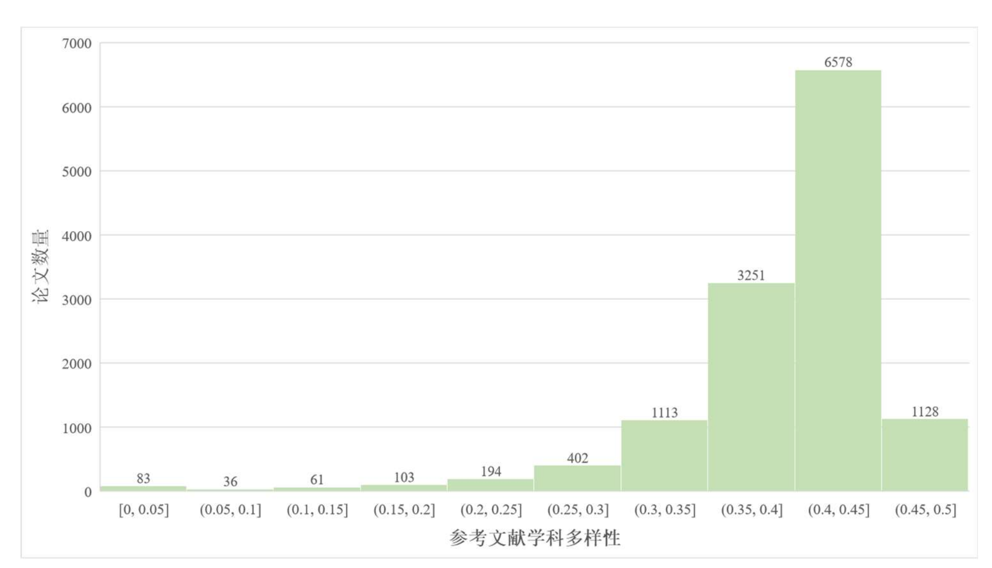


图 3 参考文献学科多样性分布情况

Figure 3 Distribution of disciplinary diversity of references

2.3 参考文献学科多样性与被引频次的关系分析

2.3.1 变量统计特征

由表 2 可见，论文参考文献学科多样性最大值为 0.467，最小值为 0，中位数为 0.410；论文被引频次的最大值为 2317，最小值为 1，方差为 2227.385，数据呈离散分布特点。为增加数据稳定性，本文对因变量作 LOG10 对数转换处理，使其满足正态分布要求以便进行回归分析。对数转换后的因变量分布见图 4，由因变量分布 P-P 图可见，因变量已呈现正态分布特征。

表 2 变量描述性统计

Table 2 Descriptive statistics of variables

	N	最小值	最大值	平均值	中位数	标准差	方差
PC5	12949	1	2317	24.52	14	47.195	2227.385
PD		0	0.630	0.368	0.368	0.006	0
CD		0.182	0.325	0.228	0.228	0.002	0
PDI		0	0.476	0.394	0.410	0.065	0.004
RN		1	292	47.25	44	25.482	649.318
AN		1	468	8.69	7	14.542	211.474
FN		0	133	2.92	2	3.352	11.238
PCN		1	7	1.99	2	0.960	0.921

TL		3	36	14.11	14	4.196	17.606
KN		0	37	4.15	5	2.478	6.139
AL		0	594	215.21	209	67.089	4500.983
PN		2	47	11.50	11	4.925	24.251
QS		2	10	7.88	8	2.297	5.275
OA		0	1	0.63	1	0.483	0.233
AS		0	1	0.01	0	0.103	0.011
IS		0	1	0.14	0	0.349	0.122
CS		0	1	0.68	1	0.468	0.219

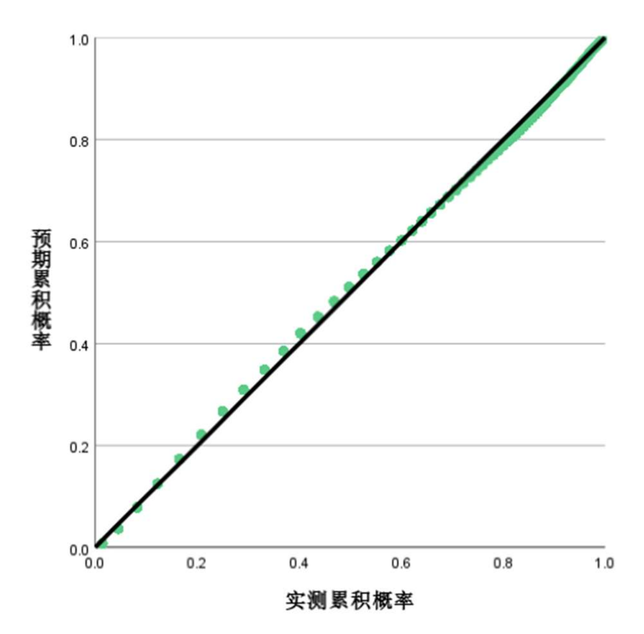


图 4 因变量正态分布检验图

Figure 4 Plot of dependent variable normal distribution test

2.3.2 变量相关性

本文使用 Spearman 相关系数检验变量相关性。由表 3 可见，参考文献学科多样性与论文被引频次呈显著正相关关系。控制变量中，除论文标题长度外，其他变量皆通过了显著性检验。除此之外，相关性结果表明自变量及控制变量间的相关系数均低于 0.7，说明变量之间不存在强相关关系，且变量间的容忍度（Tolerance）均远大于 0.1，方差膨胀因子均小于 10，说明模型不存在多重共线性问题，可以进行回归分析。

表 3 变量相关性

Table 3 Variable correlations

	PC5	**PDI**	**PD**	**CD**	**RN**	**AN**	**FN**	**PCN**	**TL**	**KN**	**AL**	**PN**	**QS**	**AS**	**IS**	**CS**	**OA**
PC5	--																
PDI	.136***	--															
PD	**-.344*****	**.161*****	--														
CD	-.062***	.287***	**.371*****														
RN	**.357*****	.303***	.035***	.043***	--												
AN	.205***	.054***	-.074***	-.048***	.075***	--											
FN	.263***	.109***	-.059***	-0.014	.265***	**.321*****	--										
PCN	.086***	.107***	.025**	0.012	.038***	-.050***	-.030***	--									
TL	-0.007	0.006	.046***	.053***	.055***	.111***	-0.014	-0.009	--								
KN	-.052***	.109***	.075***	.109***	.099***	-.118***	-.083***	-.064***	.086***	--							
AL	.072***	.095***	.030***	.018*	.234***	.103***	.037***	.040***	.194***	.112***	--						
PN	**.338*****	.185***	.020*	.029***	**.678*****	.100***	.252***	.033***	.071***	.060***	.272***	--					
QS	**.515*****	.189***	-.116***	-.035***	.387***	.226***	.323***	.135***	0.006	-.106***	.121***	.380***	--				
AS	-.063***	-.056***	.023**	0.001	-.050***	-.180***	-.094***	-.020*	-.052***	-0.007	-.084***	-.039***	-.068***	--			
IS	-.097***	-.048***	.042***	.041***	-.071***	-.281***	-.149***	0.005	-0.006	.020*	-.065***	-.059***	-.129***	.133***	--		
CS	-.124***	0.008	.090***	.082***	-.119***	-.267***	-.208***	.020*	0.010	.021*	-.040***	-.107***	-.148***	.053***	.258***	--	
OA	.100***	.051***	-.021*	0.001	.173***	.144***	.169***	-.128***	0.002	-0.013	.134***	.175***	.195***	-.036***	-.086***	-.130***	--
T		0.838	0.978	0.952	0.577	0.884	0.825	0.935	0.958	0.921	0.836	0.557	0.738	0.966	0.904	0.875	0.906
VIF		1.193	1.022	1.051	1.733	1.132	1.212	1.07	1.044	1.086	1.197	1.797	1.355	1.035	1.107	1.143	1.104

***.在 0.001 级别，相关性显著；**.在 0.01 级别，相关性显著；*.在 0.05 级别，相关性显著；T 表示容忍度；VIF 表示方差膨胀因子。

2.4 模型构建与拟合

2.4.1 模型构建

本文采用多元线性回归分析法分析论文参考文献学科多样性与论文被引频次之间的关系，将所有变量纳入回归函数建立的多元线性回归模型见公式 6：

$$\log_{10}(PC_5)=\beta_0+\beta_1 PDI+\beta_2 PD+\beta_3 CD+\beta_4 RN+\beta_5 AN+\beta_6 FN+\beta_7 PCN+\beta_8 TL+\beta_9 KN+\beta_{10} AL+\beta_{11} PN+\beta_{12} QS+\beta_{13} AS+\beta_{15} IS+\beta_{16} CS+\beta_{17} OA+\mu \quad (6)$$

模型中，β_0是截距项，β_1至β_{17}是自变量、控制变量对应的回归系数，μ为残差项。

2.4.2 回归模型拟合与结果分析

与其他方法求得的无偏估计量相比，最小二乘法（Ordinary Least Squares，OLS）收敛速度快，拟合效果好，具有估计量无偏性和最小方差性的优势[35]，因此，本文使用最小二乘法拟合回归模型。为直观体现论文参考文献学科多样性对被引频次的影响，本文采用逐步回归分析法，在模型 1 中只加入控制变量，模型 2 在模型 1 的基础上加入参考文献学科多样性，以此衡量参考文献学科多样性对论文被引频次的影响，回归结果见表 4。

由回归结果可见，模型 1 的 R^2=0.324，说明控制变量对论文被引频次的解释程度为 32.4%。在所有控制变量中，是否机构独著（IS）和是否国家独著（CS）未通过显著性检验，作者数、基金数、学科数、论文页数、期刊等级和参考文献数量正向影响论文被引频次，其他变量负向影响论文被引频次。模型 2 的 R^2 为 0.325，拟合效果优于模型 1，参考文献学科多样性在 0.001 级别显著正向影响论文被引频次，表明论文参考文献学科多样性是影响论文被引频次的因素之一。

表 4 模型拟合结果

Table 4 Regression results of model fitting

	模型 1	模型 2
变量	标准化系数	标准化系数
AN	0.100***（0.000）	0.101***（0.000）
FN	0.069***（0.001）	0.067***（0.001）
PCN	0.017*（0.004）	0.015*（0.004）
TL	-0.021**（0.001）	-0.021**（0.001）
KN	-0.028***（0.001）	-0.029***（0.001）
AL	-0.022**（0.000）	-0.023**（0.000）
PN	0.055***（0.001）	0.056***（0.001）
QS	0.429***（0.002）	0.424***（0.002）

AS	-0.022**（0.033）	-0.021**（0.033）
IS	-0.011（0.010）	-0.010（0.010）
CS	-0.010（0.008）	-0.011（0.008）
OA	-0.031***（0.007）	-0.031***（0.007）
CD	-0.051***（1.403）	-0.057***（1.424）
RN	0.128***（0.000）	0.121***（0.000）
PD	-0.056***（0.556）	-0.059***（0.558）
PDI		0.032***（0.056）
N	12949	12949
显著性	0.000	0.000
R^2	0.324	0.325
调整后 R^2	0.323	0.324

***.在 0.001 级别相关性显著；**.在 0.01 级别相关性显著；*.在 0.05 级别相关性显著；括号内为标准误。

2.5 模型稳健性检验

参考已有研究[36]，本文采用抽取子样本的方式检验模型稳健性。为排除结果的随机性，使用 RAND 函数多次抽取其中 5000 篇论文作为样本数据进行检验，结果表明模型 R^2 并无显著变化，表 5 为随机抽样数据的回归结果。由表 5 可见，回归模型整体皆显著，参考文献学科多样性对被引频次的影响关系并没有改变，表明本文构建的模型较为稳定。

表 5 模型稳健性检验

Table 5 Model robustness test

变量	标准化系数	共线性统计	
		容差	VIF
AN	0.123***（0.000）	0.856	1.168
FN	0.059***（0.002）	0.814	1.228
PCN	0.022（0.006）	0.935	1.069
TL	-0.037**（0.001）	0.956	1.046
KN	-0.019（0.002）	0.914	1.095
AL	-0.008（0.000）	0.838	1.194
PN	0.053***（0.001）	0.535	1.869
QS	0.425***（0.003）	0.736	1.359
AS	-0.031**（0.053）	0.957	1.045
IS	-0.020（0.016）	0.900	1.111
CS	-0.007（0.012）	0.881	1.135
OA	-0.024（0.012）	0.915	1.093
CD	-0.065***（2.412）	0.896	1.116
RN	0.101***（0.000）	0.551	1.814
PD	-0.051***（0.703）	0.890	1.123
PDI	0.042***（0.093）	0.839	1.192
N	5000		

显著性	0.000		
R^2	0.331		
调整后 R^2	0.329		

***.在 0.001 级别相关性显著;**.在 0.01 级别相关性显著;*.在 0.05 级别相关性显著;括号内为标准误。

3 研究结果讨论与展望/ Discussion of findings and outlook

本文设计论文参考文献学科多样性测度指标，依据测度结果对论文参考文献学科多样性与被引频次之间的关系进行了研究。经相关性分析、总体结果分析以及回归结果分析，本文发现，论文发表期刊等级是影响论文被引频次最核心的因素，其与被引频次之间呈显著正相关关系，论文发表期刊等级越高，论文获得的被引次数就越高。除期刊等级外，相较其他控制变量，论文作者数量和论文参考文献数量对被引频次的影响最为显著。

本文进一步通过逐步回归分析方法，对参考文献学科多样性和论文被引频次之间的影响关系进行了分析，研究结果表明参考文献学科多样性显著正向影响论文被引频次，加入参考文献学科多样性的回归模型拟合优度优于仅包含控制变量的模型拟合优度，表明论文参考文献学科多样性越强，论文越有可能吸引更多的引用。进一步表明融合多学科交叉的知识、思维、资源等的知识基础为论文创造了更为广泛的创新思想来源，在一定程度上能够促进论文产出价值的提升[37]，进而对论文后向引文影响力产生影响。

综上，本研究对揭示论文影响力的影响因素而言具有一定参考价值，不足之处在于当前研究仅以 3 个领域数据为样本进行分析，同时，由于数据获取难度较大，暂未将作者研究领域特征纳入控制变量中进行分析，未来将进一步扩充实证样本量，利用数智技术获取更多对论文参考文献学科多样性与被引频次之间影响关系产生作用的变量数据，从更加全面、综合、深入的视角探索论文参考文献学科多样性与被引频次之间的关系。

参考文献

[1]沈楠, 徐飞. 当今全球科研评价的二难困局及其挑战[J]. 科学学研究, 2023, 41(6):980-988. (SHEN N, XU F. The dilimma and challenge faced by current global scientific research evaluation[J]. Studies in Science of Science, 2023, 41(6):980-988.)

[2]刘嘉明, 孙建军. 参考文献跨学科性与论文学术影响力的关系研究[J]. 情报学报, 2023, 42(5):525-536. (LIU J M, SUN J J. Research on the Relationship between Interdisciplinarity of References and Academic Influence[J]. Journal of the China Society for Scientific and Technical Information, 2023, 42(5):525-536.)

[3]张丽华, 姚长青. 作者学术经验与被引频次的关系探讨[J]. 情报工程, 2023, 9(5):59-72. (ZHANG L H, YAO C Q. Research on the Relationship Between Author's Academic Experience and Citation Counts[J]. INFORMATION ANALYSIS, 2023, 9(5):59-72.)
[4]方红玲, 张亚杰, 徐自超. 第一作者和合著者的生产力、影响力与论文被引频次的相关性对比研究[J]. 中国科技期刊研究, 2024, 35(2):273-279. (FANG H L, ZHANG Y J, XU Z C. Relationship among first author and co-author productivity, influence, and cited frequency of papers[J]. Chinese Journal of Scientific and Technical Periodicals, 2024, 35(2):273-279.)
[5]Fan L X, Guo L, Wang X H, et al. Does the author's collaboration mode lead to papers' different citation impacts? An empirical analysis based on propensity score matching[J]. Journal of Informetrics, 2022, 16(4):13.
[6]Pasterkamp G, Rotmans J I, de Kleijn D V P, et al. Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles[J]. Scientometrics, 70(1):153-165.
[7]汪舒雯, 许元杰, 穆梦迪, 等. 标题中"基于"或"考虑"样式对论文被引次数的影响[J]. 图书情报知识, 2023, 40(3):139-148. (WANG S W, XU Y J, MU M D, et al. The Impact of the "Based on" or "Considered" Styles in Title on Citations of Academic Paper[J]. Documentation,Information & Knowledge, 2023, 40(3):139-148.)
[8]盛丽娜, 顾欢. 基于学科角度的论文长度与被引频次关系研究[J]. 情报杂志, 2019, 38(4):111-118+110. (SHENG L N, GU H. Exploring the Ｒelationship Between Length of Journal Articles and Citation Frequency Based on the Disciplines[J]. Journal of Intelligence, 2019, 38(4):111-118+110.)
[9]郭进京, 黄奇, 盛姝, 等. 基于倾向值匹配的即时开放获取对论文引用和下载量的影响分析[J]. 情报杂志, 2022, 41(7):166-175. (GUO J J, HUANG Q, SHENG S, et al. Analysis of the Immediate Open Access's Impact on the Citations and Downloads of Papers Based on Propensity Score Matching[J]. Journal of Intelligence, 2022, 41(7):166-175.)
[10]Arthur, W B. The structure of invention. Research Policy, 2007, 36(2), 274–287.
[11]何传启, 张风. 知识创新[M]. 北京:经济管理出版社, 2001. (HE C Q, ZHANG F. Knowledge innovation[M]. Beijing:Economic Management Press, 2001.)
[12]李正风. 科学知识生产方式及其演变[M]. 北京:清华大学出版社, 2006. (LI Z F. Scientific Knowledge Production and its Evolution[M]. Beijing: Tsinghua University Press, 2006.)
[13]姜磊, 林德明. 参考文献对论文被引频次的影响研究[J]. 科研管理, 2015, 36(1):121-126. (JIANG L, LIN D M. Ｒesearch on the impact of reference on citation[J]. Science Research Management, 2015, 36(1):121-126.)
[14]葛赵青, 苗凌, 赵大良, 等. 科技期刊参考文献数量与部分引证指标的定量关系初探[J]. 编辑学报, 2015, 27(5):423-425. (GE Z Q, MIAO L, ZHAO D L, et al. Preliminary analysis of quantitative relationship between the number of references and citation indexes for Chinese sci-tech journals[J]. Acta Editologica, 2015, 27(5):423-425.)
[15]Bornmann L, Schier H, Marx W, et al. What factors determine citation counts of publications in chemistry besides their quality?[J]. Journal of Informetrics, 2012, 6(1):11-18.
[16]张洋, 常赵鑫, 龙秋爽, 等.学科布局多样性与学科竞争力的关系研究——以图书情报与档案管理学科为例[J]. 情报理论与实践, 2023, 46(1):72-80. (ZHANG Y, CHANG Z X, LONG Q S, et al. Research on the Relationship Between Discipline Layout Diversity and Discipline Competitiveness:An example from Library and Information Science[J]. Information Studies:Theory & Application, 2023, 46(1):72-80.)
[17]张雪, 张志强. 学科交叉研究系统综述[J]. 图书情报工作, 2020, 64(14):112-125. (ZHANG X, ZHANG Z Q. Review on Interdisciplinary Research[J]. Library and Information Service, 2020, 64(14):112-125.)
[18]顾秀丽, 黄颖, 孙蓓蓓, 等. 图书情报领域中的交叉科学研究:进展与展望[J]. 情报学报, 2020, 39(5):478-491. (GU X L, HUANG Y, SUN B B, et al. Interdisciplinary Research in Information Science and Library Science: Progress and Prospects[J]. Journal of the China Society for Scientific and Technical Information, 2020, 39(5):478-491.)
[19]Stirling A. A general framework for analysing diversity in science, technology and society[J]. Journal of the

Royal Society Interface, 2007, 4(15):707-719.
[20]Li J, Yang X, Lu X L, et al. Making journals more international: Language subject differences and impact performance[J]. Learned Publishing, 2023, 36(4):596-618.
[21]梁春慧, 孙艳, 万跃华. 高被引论文的参考文献特征研究——以化学领域为例的实证分析[J]. 科技与出版, 2014(7):119-122. (LIANG C H, SUN Y, WAN Y H. A study of reference characteristics of highly cited papers - an empirical analysis with the example of chemistry field[J]. Science-Technology & Publication, 2014(7):119-122.)
[22]Lovaglia J M. Predicting citations to journal articles:the Ideal number of references[J]. Am Sociol, 1991, 22:49-64.
[23]GREGORY D W, PETER K J, TATIANA O S. Hot topics and popular papers in evolutionary psychology: analyses of title words and citation counts in evolution and human behavior, 1979-2008[J]. Evolutionary psychology, 2009, 7(3):348-362.
[24]刘小慧, 朱曼曼. 科研论文颠覆性与参考文献年龄之间的关系研究[J]. 情报理论与实践, 2023, 46(7):60-66. (LIU X H, ZHU M M. Research on the Relationship between the Disruption of Scientific Papers and the Age of References[J]. Information Studies:Theory & Application, 2023, 46(7):60-66.)
[25]Ardito L, Petruzzelli A M, Dezi L, et al. The influence of inbound open innovation on ambidexterity performance: Does it pay to source knowledge from supply chain stakeholders?[J] Journal of Business Research, 2020, 119:321-329.
[26]Trantopoulos K, Von Krogh G, Wallin M W, et al. EXTERNAL KNOWLEDGE AND INFORMATION TECHNOLOGY: IMPLICATIONS FOR PROCESS INNOVATION PERFORMANCE[J]. Mis Quarterly, 2017, 41(1):287-+.
[27]康旭东, 徐庆富, 张春博. 期刊单元下参考文献与论文被引频次间的关系——以 Web of Science 凝聚态物理学科为例[J]. 中国科技期刊研究, 2018, 29(6):619-626. (KANG X D, XU Q F, ZHANG C B. Relationship between reference and paper citation based on journal unit: A case study of condensed matter physics in Web of Science[J]. Chinese Journal of Scientific and Technical Periodicals, 2018, 29(6):619-626.)
[28]董克, 陈晓萍, 吴佳纯. 科研论文创新性与引文影响力相关性研究——基于语义视角的测度[J]. 情报理论与实践, 2023, 46(10):24-31. (DONG K, CHEN X P, WU J C. Research on the Correlation between Creativity and Citation Impact of Scientific Research Paper: Measurement from Semantic Perspective[J]. Information Studies:Theory & Application, 2023, 46(10):24-31)
[29]Wu L F, Wang D S, Evans J A. Large teams develop and small teams disrupt science and technology[J]. Nature, 2019, 566(7744):378-+.
[30]刘雪立, 申蓝, 郭佳, 等. 利用论文被引频次基准线遴选学者代表作的方法——以河南省学者论文为例[J]. 中国科技期刊研究, 2020, 31(8):941-947. (LIU X L, SHEN L, GUO J, et al. Method for selecting scholars' representative works by baselines of citations: Based on the articles of Henan Provincial scholars[J]. Chinese Journal of Scientific and Technical Periodicals, 2020, 31(8):941-947.)
[31]刘仲林, 赵晓春. 跨学科研究:科学原创性成果的动力之源——以百年诺贝尔生理学和医学奖获奖成果为例[J]. 科学技术哲学研究, 2005, 6:108. (LIU Z L, ZHAO X C. Interdisciplinary Research: A Powerhouse of Scientific Originality--An Example of 100 Years of Nobel Prize in Physiology and Medicine Winning Results[J]. Studies in Philosophy of Science and Technology, 2005, 6:108.)
[32]Lin Z H, Yin Y, Liu L, et al. SciSciNet:A large-scale open data lake for the science of science research[J]. Scientific Data, 2023, 10(1):22.
[33] Bornmann L, Tekles A. Disruption index depends on length of citation window[J]. Profesional De La Informacion, 2019, 28(2):2.
[34]张雪, 刘昊, 张志强. 不同合作模式下的学科交叉程度与文献学术影响力关系研究[J]. 情报杂志, 2021, 40(8):164-172. (ZHANG X, LIU H, ZHANG Z Q. Research on the Relationship Between Interdisciplinarity and

Academic Influence of Literature under Different Collaboration Modes[J]. Journal of Intelligence, 2021, 40(8):164-172.)

[35]林祥都编著，经济计量学理论与方法[M]. 西安，西安电子科技大学出版社,1993. (LIN X D, Econometrics Theory and Methods[M]. XIAN, Xi'an Electronic Science and Technology University Press, 1993.)

[36]徐孝婷，杨梦晴，宋小康. 在线健康社区中医生口碑对患者选择的影响研究——以好大夫在线为例[J]. 现代情报, 2019, 39(8):20-28+36. (XU X T, YANG M Q, SONG X K. Exploring the Impact of Physicians'Word of Mouth on Patients'Selection in Online Health Community——Taking the Website of www．haodf．com as an Example[J]. Journal of Modern Information, 2019, 39(8):20-28+36.)

[37]罗生全. 学科交叉的源动力与创新发展机制[J]. 南京社会科学, 2022(9):140-146. (LUO S Q. Source Power and Innovative Development Mechanism of Interdisciplinary[J]. Nanjing Journal of Social Sciences, 2022(9):140-146.)

(通信作者：杨雪 Email:yangxue2115@163.com)

作者贡献说明/**Author contributions**：

杨雪：选题，研究设计，数据收集与处理，论文撰写与修改。

Research on the Relationship between the Diversity of Academic Disciplines in References and the Citations

Yang Xue

School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [Purpose/Significance] The paper constructs the index of the discipline diversity of the paper reference, and explores the relationship between the discipline diversity of the paper reference and the citations. **[Method/Process]** The reference discipline diversity measurement index was designed based on the Rao-Stirling diversity index, and 15 variables including the reference discipline diversity of the paper were taken as the independent variable, the citation of the paper as the dependent variable, and the journal rank as the control variable. The regression model was generated by fitting the ordinary least square method. This paper discusses the relationship between the subject diversity of the paper references and the citation frequency. **[Result/Conclusion]** The index of reference diversity constructed by this research can reflect the degree of multidisciplinary knowledge integration of the paper. The diversity of the reference disciplines positively affects the citation frequency of the paper. The stronger the diversity of the reference disciplines, the higher the citation of the paper.

Keywords: science and technology evaluation references disciplinary diversity interdisciplinarity scientific academic impact

Authors: Yang Xue, M.S. candidate, School of Information Management, Sun Yat-sen University, yangxue2115@163.com.